10. Quantum physics

Draft of 14 April, 2005

Question: What do all the following phenomena have in common?

lasers
solar cells
transistors
computer circuitry (integrated circuits)
digital camera CCDs
superconductors

Answer: They all make use of "quantum" physics that were discovered in the 20th century. Without an understanding of quantum physics, these things could not be understood, designed, or used effectively.

In fact, it could be said that most of what we call *high tech* is founded on quantum physics. What is quantum physics?

Quantum physics is basically the recognition that there is less difference between waves and particles then we thought. The key insights are these:

1. everything we thought was a particle is actually a wave (or wave packet)
2. waves gain or lose energy only in "quantized amounts" (the quantum leap)
3. when a particle is detected, it's wave will usually change suddenly into a new wave

Those simple statements are, of course, tricky to understand. But they lead to all the mysterious behavior of quantum physics, including not only the list given above, but also the famous Heisenberg *uncertainty principle*, a result that has had major impact on philosophy and even the way that people unfamiliar with physics think about life. The uncertainty principle that showed that physics was theoretically incapable of predicting everything in the future.

But before we get into a discussion of this, let's examine the features of quantum physics. We begin by looking closely at the consequences of the fact that an electron is a wave.

Electron waves

All electrons, protons, indeed all particles, behave in the same quantum manner: when detected, they have some features that make them look like particles (they deliver their energy in bursts), but between detections, they travel like waves.

Think about an atom. The electron orbits the nucleus in much the same way that the Earth orbits the Sun – except that it is held in its orbit by electric force rather than by gravity. But now here is something new. In quantum physics, we must think of the electron as a wave. The frequency of the wave is given by the first key equation of quantum physics, often called the Einstein equation. It relates the energy of the wave to its frequency:

Einstein Equation:

$$E = h\,f$$

The constant h is called "**Planck's constant**." Planck had found this constant when he studied the behavior of light. It's value is $h = 6\text{x}^{-34}$ joules per Hz.

For a typical atomic energy of about 1 eV (= $1.6\text{x}10^{-19}$ joules), this frequency is very high:

$$f = E/h$$

$$= 1.6\text{x}10^{-19} / 6\text{x}10^{-34}$$

$$= 2.7\text{x}10^{14} \text{ cycles per sec (i.e. Hz)}$$

$$= 270{,}000{,}000{,}000{,}000 \text{ Hz}$$

So electron waves oscillate with an extremely high frequency. That's why you never noticed it. It oscillates so fast that you can't perceive the oscillation. Remember when you heard a music tone? It too was an oscillation, but except at the very low frequencies, you never perceived it as such. An electron oscillation is similar. With the right instruments, we can now measure these frequencies (by seeing beats with other oscillations), but we don't sense them directly.

electron waves in atoms

Now here is the way that the wave behavior really makes a difference: in an atom, the wavepacket is usually longer than the circumference. That means that when orbiting the atoms, the electron wave runs into its own tail. Since it is a wave, it can actually cancel itself out.

If it cancels itself out, then there is no electron. So if the electron is orbiting the nucleus, the only possible orbits are those where, after an orbit, the wave doesn't cancel. That means that only certain frequencies are allowed.

According to the Einstein equation, frequency is related to energy. So if only certain frequencies are allowed, then only certain *energies* are allowed. This is sometimes stated by saying that the electron energy is *quantized*.

The allowed kinetic energies for the hydrogen atom – the electron energies of the waves that circle the atom -- have been calculated, and they match the calculated energies for which the orbiting wave doesn't cancel itself. These *allowed* energies include the following: 13.6 eV, 3.4 eV, 1.5 eV, 0.85 eV. These are the energies that don't lead to self-cancellation. There are other allowed energies too. These kinetic energies can all be described by a simple formula that was discovered by Niels Bohr:

$$KE = 13.6/n^2$$

where n is an integer (n = 1, 2, 3, …). Different values of n give the energies in the list. For n = 1, we get 13.6 eV. For n = 2, we get 3.4 eV, and so on for all the energies possible. Notice that there are an infinite number of allowed energies, but there are many energies (those inbetween the allowed one) that are forbidden. No electron can orbit the electron unless its energy matches one from the Bohr formula.

Although energies of orbiting electrons are always quantized, an electron that is moving through empty space in a straight line can have any kinetic energy. The quantization of energy came about for the hydrogen atom because the electron was moving in an orbit and must not cancel itself. So not all energies are quantized.

Fingerprints of different atoms

When one hydrogen molecule collides with another (as it does in a gas) it can knock the electron from one allowed orbit to another one that has a greater energy. Put another way, a collision can alter the energy of an atom. If it increases the energy of the electron, we say that the atom is *excited*, and that the electron is in an excited orbit. The electron can lose this energy and fall back into the original orbit by radiating an electromagnetic wave that carries off the energy difference.

In our new language of quantum physics, we say that the electron changes its orbit (a quantum jump) and emits a light wave. If the energy *difference* between the two orbits is *E*, then *E* is the energy carried by the light wave.

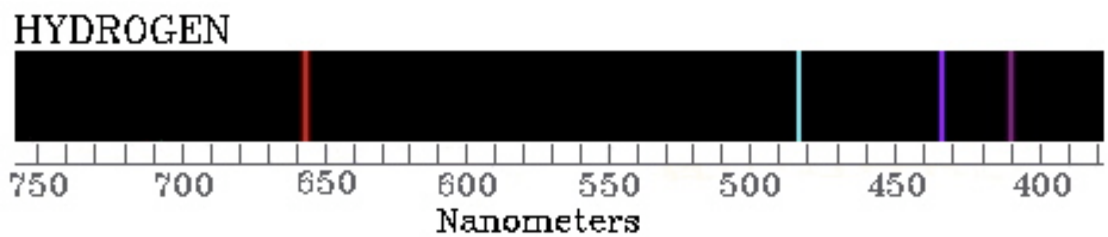
The frequency of this emitted light wave is given by the Einstein equation

$$E = h\,f$$

Note that we have now used this equation twice, once for the electron (to put it in an allowed orbit) and once for the light that was emitted.

This logic implies that for a given atom, the frequency of the emitted light has only certain values – those that correspond to the differences in the quantized energy levels. But frequency is color, so that means that the color of the emitted light is also quantized. Only certain colors are allowed.

The quantization of color can be seen by putting the light from hot hydrogen gas through a prism to separate out the colors. Instead of a continuous spectrum (as we got from the sun) we see that only a few colors are present, as in the figure below. The "Nanometer" scale refers to the wavelength of the light.



(borrowed from http://www.nhn.ou.edu/~kieran/reuhome/vizqm/figs)

A hot hydrogen gas will have lots of collisions, and so will always emit the pattern of spectral lines shown above. If you see the pattern seen above, then you know the gas is hydrogen.

For comparison, the figure below shows the colors of the rainbow – the colors that you get when you shine sunlight (white light) on a prism.

In the both images above, the colors were often spread out in the vertical direction. For the hydrogen case, that made the color appear as lines. For this historic reason the quantized colors are often still referred to as "spectral lines."

The figure below shows the spectra from hydrogen and helium compared.

Photo courtesy NASA
Hydrogen spectrum

Photo courtesy NASA
Helium spectrum

Why is the spectrum of helium different? The reason is that helium has two protons in its nucleus, so the electron has to orbit at a higher speed to balance that attraction. Higher energy means higher frequency. That means that the orbits that don't cancel will be different.

Think of these patterns as "fingerprints." They are so easy to tell apart that even if the gases hydrogen and helium were mixed, we could tell how much hydrogen and how much helium was present.

The origin of such spectral lines was once considered a mystery. They were finally understood only through the creation of the theory of quantum physics. Einstein came up with his equation in 1905, and Bohr came up with his formula for the colors of the hydrogen atom in 1913. Now the existence of these lines is no longer considered to be a mystery. The fingerprints provide an incredibly powerful way to identify elements and molecules. Measurement of spectra such as these has allowed us to determine the gases on the surface of distant stars, and even to measure the gaseous emissions from smokestacks that may be emitting illegal pollution.

The fact that electrons can only change energy by quantum leaps makes possible one of the most fascinating inventions of the 20th century: the laser.

Photons

When Einstein first came up with his equation

$$E = h\,f$$

he wasn't applying it to electrons (although it does apply to them). Rather, he was thinking about light. He had concluded (based on the observation of such absorption by others) that when a light wave is emitted or absorbed, the change in the energy of the wave is always quantized, by an amount given by the equation.

Put another way, waves cannot have any arbitrary amount of energy. The possible energy it has must be a multiple of $h f$. So, for example, the allowed energies for a light wave always can be written as

$$E = N h f$$

where N is a whole number ($N = 1$ or 2 or 3 …).

What does N mean? It has a simple interpretation. We say that N is the number of ***photons*** present.

Think of a photon as a quantum of light. If you have a light wave, it can be absorbed – but only in an amount equal to an integral number of photons.

If you know the frequency of a light wave, then you know the energy of the photons – from the Einstein equation. A very bright light has many photons present. A dim light has only a few.

In the language of quantum physics, we say that light is a quantized wave packet. Every time it gains or loses energy, it does so one photon at a time. It behaves like a wave, but it also behaves like a collection of photons. This is sometimes called the wave-particle *duality* of quantum physics.

Laser – a quantum chain reaction

Lasers are used to burn holes in metal, to send information at enormously high rates over fibers, to read supermarket labels, to measure the exact shape of irreplaceable sculptures, to give spectacular light shows, as convenient pointers, to make holograms, and to find the distance to a remote object (including the moon). Future uses may include the triggering of controlled nuclear fusion, and in shooting down military airplanes and ballistic missiles.

Lasers work on the principle of *stimulated emission*, another effect predicted by Einstein. Recall that an electron emits light when it makes a quantum leap from one allowed energy orbit to another. If there already is light present, then that light could reinforce the emitted light. When that happens, the the probability of emitting the new energy is increased.

Let me say the same thing, but in the new language of quantum physics. When one photon is present, that makes the probability higher that another photon will be emitted. This enhanced probability is what is called stimulated emission. A key and important feature of stimulated emission is that the second photon has the same frequency and direction as the photon that stimulated it to be emitted.

The word ***laser*** is an acronym for “light amplification by stimulated emission of radiation.” You should learn the words that make up this acronym.

In a typical laser there are a huge number of atoms present (e.g. 10^{20}). The first step is to make sure the electrons in many of these atoms are in an "excited state" – which means they are in an orbit with extra energy. This is sometimes done by hitting these atoms with another source of light, that gives the electrons more energy.

One atom emits a photon, and that stimulates more photon emission. Mirrors, one at each end, reflect the photons back and forth to make for even more stimulated emission. An avalanche develops. The two photons trigger another two, and these four trigger four, then eight, sixteen, etc. The reaction stops only when all of the atoms are de-excited. The "a" in laser could also stand for "*avalanche*", but it doesn't.[1]

The first experiments with this effect used microwaves instead of visible light, and the device was called a “maser” (for microwave emission by stimulated emission of radiation). Lasers now work in the infrared, the ultraviolet, and work has been done to try to get them to work with x-rays. The principle for all these is the same: a photon avalanche, or equivalently, a photon chain reaction.

As with the nuclear bomb, the chain reaction can happen very quickly. When this occurs, the pulse of light can be extremely powerful, although very short. Such lasers are called pulsed lasers, and they are the most powerful ones. These are the kinds that are being used at our national laboratories in attempts to ignite nuclear fusion without using a fission primary. They are also the kind that the military is developing for laser weapons.

However, it is also possible to operate the laser in a continuous manner in which the light output remains constant. (That makes it analogous to the sustained chain reaction in a nuclear reactor.) To do this, we must continuously excite new atoms at the same rate that they are emitting. This is done in a gas laser by sending an electric current continuously through the gas. Continuous lasers are used for laser communications (through fiber optics), and for measurement (range finding and leveling). Continuous lasers are also used for laser pointers and supermarket label readers.

The laser has two important properties that make it unlike the other chain reactions we studied, and contribute enormously to its value. I mentioned these earlier. They are:

-- The emitted photons all have identical frequency
-- The emitted photons all have identical direction

Identical frequencies means that the light is only one color, i.e. it is monochromatic. This is the feature that makes lasers really valuable for communications. Information is

---

[1] According to Charles Townes, the original inventor of the laser, they briefly considered an alternative name: electromagnetic radiation by stimulated emission of radiation. If they had kept those words, the acronym today would be *eraser*.

carried by a laser beam by modulating it, i.e. by changing its amplitude. You want the original beam to be as constant as possible. If there are multiple frequencies in the beam, then beats between these frequencies will give a false modulation.

You'll sometimes hear that the light from a laser is *coherent*. Coherent is a fancy word meaning that only one frequency is present – or at least that the range of frequencies present is very small.[2]

The fact that the emitted photons have identical direction is more important than you might guess. It means that the beam comes out of the laser with all the light parallel, well collimated. That's why a laser beam doesn't seem to spread very much, unlike a flashlight beam or headlight of a car. Even sunlight has light coming from different directions: since the sun covers about a half angular degree in the sky, light from different parts comes at slightly different directions. But laser beams are different. Of course, they will spread a little, since they are waves. But that spreading angle can be tiny, since the wavelength is so short.[3] Sometimes it is necessary to spread a laser beam, for example, if you want it to illuminate a hologram. You can do this by passing it through a lens or bouncing it off a curved surface. But the light originally produced is very well collimated.

laser measurements

The collimation of the narrow laser beam makes them useful for measurement that otherwise might be difficult. A pulsed laser beam can directed at a distance object, and the bright spot on the object can be observed in a telescope and have its time to return accurately measured. That time then gives the distance to the object. That is the basis for laser rangefinding. If you measure distance for many different directions, you get a record of the entire shape of an object. Lasers have been used in this way to measure the changing shape of volcanoes, the interiors of buildings and caves, and historic structures such as the Roman Coliseum. Laser scanners, based on similar principles, are now being used for very detailed measurements of the shapes of objects, including valuable and irreplaceable sculptures.

At construction sites, a laser beam can be made level, and that beam then placed across the entire construction site to make the structures level with each other. Lasers are sometimes lined up with boundaries of property, to see what objects are in the property

---

[2] Optional: the "coherence time" is equal to one divided by the bandwidth, i.e. the difference in the maximum frequency present minus the minimum. If the bandwidth is small, then the coherence time is very long.

[3] The equation for spreading is the one we discussed in chapter 8: B = (L/D) R. For a beam of diameter D = 5 mm = 5 x10^{-3} meters, and L = 0.5 microns = 5 x10^{-7} meters, the (L/D) = 10^{-4}. So in R = 100 meters, it will spread only 1 cm. In 1 km, it will spread to a size of 10 cm. If you see a laser light show, by the time the end of the beam is 1 km away, it still looks pretty small.

and which are outside. They are particularly useful for hilly land where the surveyor could not lay out a string. They can be used to show construction workers exactly where to lay foundations or columns.

An intriguing use of lasers for measurement was done for the movie “The Two Towers” (part of the Lord of the Rings trilogy). An actor (Andy Serkis) climbed down cliffs, walked on all fours, moved in complex ways – and the positions of his shoulders, head, hands, and other parts of his body were measured using lasers to detect corner reflectors that had been attached to them. A computer then generated a new image, completely computer-generated, of an imaginary creature named Gollum.

supermarket lasers for bar-code reading

Supermarket lasers emit a very narrow (less than a millimeter across) single color laser beam, and they move its direction rapidly (they scan it) in a complex pattern. The supermarket checker puts the part of the item for sale called the "bar code" in this scanning laser beam.

In addition to the laser, there is a detector that looks at reflected light. The detector is designed to measure only light that matches the frequency of the laser. (They use a filter to eliminate all other light.)

When the beam scans across the bar code on the product, the reflected light changes rapidly, matching the dark and light spots on the code. The detector notices this rapid blinking, and records the pattern. From the pattern, it can look up the price in a catalogue, or just record the fact that the item was purchased. The narrowness of the laser beam is important for being able to record the narrow pattern.

It turns out that the easiest way to point the laser beam is not by moving the laser, but by bouncing it off a spinning hologram. Different parts of the hologram point the beam in different directions. So supermarket scanners make use of two very high-tech devices: lasers and holograms.

laser cleaning

Lasers are being used to clean old and valuable statues, without damaging the surface. To do this, they take advantage of their ability to deliver high power for very short pulses. A laser pulse delivered to the surface can cause very intense heating, enough to vaporize soot and oil, but if the pulse is very short (typically a laser is used with a pulse that lasts only a few nanoseconds), then it is only a very thin layer of the statue that is heated. The image below shows a statue that was cleaned in this manner.

Ancient sculpture, before and after laser cleaning
(image borrowed from www.buildingconservation.com/articles/laser/laser.htm)

Needless to say, when something is developed for scientific or artistic reasons, someone will figure out how to make money from the process. Laser cleaning and whitening of teeth is already being practiced in the United States, and dentists are looking seriously into the use of lasers for removal of dental caries and other medical procedures.

laser weapons

Ever since the laser was invented, the military has looked for potential weapons applications. Lasers can deliver a lot of energy at the speed of light. This application was limited, at first, by the huge size and enormous weight of the very energetic lasers that were available. But recent development of portable lasers using carbon-dioxide have revived interest. Such lasers could, in principle, be carried on airplanes.

Note that the problem with weight is not from having to point the laser. The laser does not have to be turned. The laser can be pointed by moving a mirror, and that can be light in weight.

Lasers have already been used to shoot down small, unmanned aircraft called drones. Lasers are frequently mentioned for their possible application for shooting down missiles. This is a situation in which speed is needed, since the missiles travel fast, and you need to destroy them before they hit their target.

The laser beam does its damage by depositing heat on the surface. If the surface is moving, then the laser beam must follow the same spot. A potential problem with such systems is that laser beams can be reflected. If the target has a mirror-like surface, then little light is absorbed, and the laser "weapon" does no damage. It is also difficult to heat the surface of a missile if the missile is rotating, since the spot exposed to the beam is constantly changing. For this reason, laser weapons may have a limited future.

A more serious application for the military is as an anti-satellite weapon. A laser can deliver a substantial amount of energy to a satellite over a period of a few minutes, and the satellite has no way to lose that energy except by heating up and radiating it. Most satellites are severely damaged if they are heated only a few tens of degrees Celsius.

Laser eye safety

Lasers can be dangerous to eyes for several reasons. The simplest, of course, is that they have high power that can be concentrated on a small spot, and the eye is delicate. But there are other reasons. The light from lasers is usually highly collimated, and parallel light is focused by the lens of the eye on the retina. Even relatively weak laser light can become intense when focused in this way. If the laser light is in the infrared or some other invisible wavelength, then the victim may not even know his eye is being damaged. For these reasons, people who work in or visit laser laboratories are usually required to wear special goggles that block out all light of the laser frequency, while allowing other light through.

laser eye surgery

Lasers have found an important use in surgery, particular for the eye. A broad laser beam can enter the eye, and be focused on a tiny spot. Because the power of the beam is spread out everywhere except at the focus, there is not much heating except at the target spot. Perhaps the most exciting application of this technique is to “weld” a detached retina to the back surface of the eye. This procedure is now common, and it has prevented thousands of people from going blind. The irony here is that the very aspect of a laser that makes it dangerous is the one that makes it useful for medicine. (Of course, the same is also true of a knife.)

Lasers are also used to cut away parts of the cornea, to reshape it so that it focuses better on the retina. This has given “normal” eyesight to people who otherwise would have to wear glasses or contact lenses. Such surgery can be done in a few minutes in a doctor’s office. The most popular kind of this surgery is called “Lasik” for laser-assisted in-situ keratomileusis. In this procedure, knife is used to open a flap on the cornea, and then a laser is used to vaporize and remove portions of the underlying cornea. the flap is put back in place, and the patient walks out of the doctor’s office. The patient can see immediately. The eye takes several days for its initial healing, and is not completely normal for several months.

(Of course, this kind of surgery is not able to cure the loss of accommodation that occurs with age.)

Lasers are also used for other kinds of eye surgery. One of these is to stop the bleeding of blood vessels in the retina that lead to an illness called "macular degeneration." The heat from the laser, delivered precisely to the blood vessel, can cause the blood to clot and seal the leak. The effect on the blood is called laser photocoagulation. Untreated, macular degeneration leads to loss of most vision.

Lasers are also being tried for other types of surgery. The highly focused beam can cut a very small region, and the heat automatically cauterizes the cut flesh (i.e. stops it from bleeding). Moreover, there is no need to sterilize a laser beam, in the way that people have to sterilize knives.

Solar cells and digital cameras

When light hits a solar cell, each photon has enough energy to knock an electron out of an atom. When it does this, the liberated electron can be used to create an electric current. This is the basis of the operation of solar cells. They turn the energy of sunlight into electric current.

Solar cells may very well become a major source of electric power in the future. Present day solar cells produce power at about three times the cost of power from natural gas, oil, or coal. An economist might say that this estimate is true only if you don't include environmental costs, such as the damage done to the environment and possible global warming. If you include those cost in, then solar cells may already be cheaper than fossil fuels.

The future of solar cells can become brighter if their cost can be brought down, or if the cost of fossil fuels increases significantly. A promising technology for cheap solar cells uses *amorphous* silicon; check the web for details.

Digital cameras work in exactly the same way as solar cells: they turn photons into electricity. In a digital camera the light is focused on an array of photocells; there is one photocell for each pixel, i.e. for each picture element. An 8 megapixel camera has 8 million of these photocells. When a photon hits any one of these cells, the emitted electron results in an electric current, and that can be read by the small computer that is part of every digital camera. There are two kinds of photocell arrays in common use. The first is called the CCD array (for "charge coupled device"), and the other is the Mosfet (for "metal oxide semiconductor field effect transistor" – you don't have to know this). The differences between these are not important for this course; they key thing to know is that they are arrays of photodetectors.

Some of the first digital cameras ever used were aboard United States spy satellites. They could take the image and use radio signals to send the image back down to the ground. At that time, even the fact that this could be done was highly classified.

image intensifiers

The human eye and brain aren't sensitive enough to sense the light from a single photon. Neither is a typical digital camera. But there is an electronic device that can, called an "image intensifier." It works on a principle very similar to that in photocells – using a photon to create an electric current. But in an image intensifier, the light is used to create an immediate image that is bright enough for the eye to see.

A modern image intensifier consists of a large number of narrow tubes, tightly packed together in a configuration called a “multichannel plate.” When a photon enters the end of one of the thin tubes, it typically hits the wall of the tube before going very far. Visible photons, even in very dim light, have an energy of 2.4 eV, and that is enough to knock an electron off the surface. (This is called the “photoelectric effect”, and it is what inspired Einstein to create his photon equation.) An electric field accelerates the electron down the tube, and it soon crashes into the side, and knocks out additional electrons. This process continues as an avalanche, and by the time all the electrons reach the end of the tube, the electron signal can be very large, a billion electrons or more. These electrons hit a phosphor, and if there are enough of them, they make a bright spot.

The entire stack of tubes can be placed at the back of a camera. The photons hit the multichannel plate instead of film, and they trigger the electron avalanche that eventually emerges from the end of the tube hits a piece of glass coated with a phosphor. If a photon entered the front end of the stack, then there will be a bright spot at the other end – bright enough for a human to see. Multichannel plates are used in most of the inexpensive image intensifiers that can be purchased on the web.

The original image intensifiers were used to make "starlight scopes." These were camera like systems that could be worn on a person's head (like attached binoculars) and used to see things in dark places. But they do require some light – that's why the word "starlight" appeared in the names. They are one of the technologies used in "night vision." Contrast them with IR night vision scopes. IR imagers can work in total darkness, using only the IR light emitted by the warmth of the person or object. Starlight scopes require a little bit of visible light to be present, and then they only amplify that.

Xerox^{tm} machines and laser printers

The Xerox machine (the generic term is “photocopier”) takes advantage of the unusual properties of the element selenium. If you put charge on a selenium surface, the charge stays there; selenium is *not* a good conductor of electricity. However, if you shine light on a region of the selenium surface, the energy of the absorbed photons is sufficient to eject the charge.

If the selenium is then exposed to a cloud of carbon soot, the soot will be attracted by the electric field to the charged regions. The result is that every where that light hit the surface stays clean, and all the places where no light hit, get sooty.

Once the sooty selenium is ready, a piece a paper is brought into contact with it, and it picks up the carbon. That dirty paper becomes the Xerox copy. The soot contains a binding material, and when the paper is heated (on the way out of the machine) the soot is permanently bound to the surface.

If the paper gets stuck before it is heated, and you have to open the machine to extract it, you'll find that the soot doesn't stick, and your hands and anything else that touches the paper become dirty.

A laser printer is a Xerox machine in which a laser is used to expose the selenium instead of an optical image. The laser scans across the surface with a fine beam whose brightness varies in just the way needed to produce the image.

Compact disks and DVDs

CDs and DVDs make use of the fact that a laser beam can be focused to a small spot. The compact disk has music recorded on a thin layer of aluminum buried inside the plastic. The music has been recorded with small bumps and "lands" between the bumps, about 0.5 microns wide and about 1 micron long. Each spot represents a 0 or a 1, and the reflected intensity is measured to read the pattern. The light shines on only one bit at a time. The CD is spun, about 1.4 million such bits pass the focused laser beam every second. The CD player can distinguish between 0s and 1s from the amount of scattered light that comes back, at this megahertz rate.

Because the laser beam can be focused, it is possible to record even more information on a disk by having several layers on one disk. This method, along with smaller bump size, is used for advanced DVD recording to enable them to record long movies. The outermost layers have to be partially transparent so that some of the light passes through it to the deeper layer. To read the bumps on one of the two layers, the light is focused on it. Any light reaching wrong layer will be out of focus. Because the spot is broad, it takes a longer time for the bumps and lands to pass under it, so the reflected pulses are longer in duration. These longer pulses (from the unwanted layer) can be eliminated by the electronics. DVD players typically have four layers (they look at two from the top and two from the underside) to store all the information for a movie. Combine that with the smaller bump size, and an advanced DVD can hold seven times as much information as a CD. The first DVD player was marketed in 1997.

Recordable CDs and DVDs don't use aluminum as the reflector, but instead they use a chemical whose reflection properties can be altered by heat. When you record on one of these disks, you heat up a tiny region enough to alter it. When you later "play" the disk, the laser is low intensity, too little to alter the spot, but the amount of reflected light shows whether the spot is shiny or dull. A detector that measures the reflected light then feeds that signal into the computer. The computer turns the 0s and 1s into an audio (or, for DVDs, visual) output.

Gamma rays – and x-rays, again

Let's use the Einstein equation to compare the energy of a gamma-ray photon to that of a visible photon.

In Chapter 9 (invisible light) we said that the frequency of a typical gamma ray was about 3×10^{21} Hz. Visible light had a frequency of about 6×10^{14} Hz. So the gamma frequency is higher by a factor of $3 \times 10^{21}/6 \times 10^{14}$ = 5 million.

That means that the gamma energy is also 5 million times greater than the visible photon energy. A typical visible photon has an energy of 2.5 eV. A typical gamma ray energy will be 5 million times greater, about 10 MeV. That's enough energy to break a deuterium nucleus into its constituent proton and neutron. Radioactive decays, which release energies typically in the MeV range, often emit gamma rays.

Let's do a similar thing for x-rays. According to the Chapter 9 table, typical x-rays have frequencies of about 10^{19} Hz. That is about 20,000 times greater than the visible light frequency. According the Einstein photon equation, the energy for x-ray photons will be 20,000 times greater. Since visible light is typically 2.5 eV per photon, that means that an x-ray photon will have energy 20,000 x 2.5 = 50,000 eV = 50 keV.

X-rays are often made by taking electrons with kinetic energy of 50 to 100 keV and slamming them into tungsten or other metals. This calculation shows that most of the energy of the electron comes out as the energy of the x-ray photon.

Suppose we have a beam of visible photons with a total energy of 10 MeV. Since each photon has only 2.5 eV, that means that there are 4 millon photons present. Each photon will lose 2.5 eV when it hits a molecule. But a gamma ray beam with the same energy will have only one gamma ray photon. When it is absorbed, it will depose all its energy at one place. Gamma rays will never be absorbed bit by bit.[4] In that sense, they appear to be more like a particle than do visible photons.

It is this large energy per photon that gives gamma rays and xrays many of their distinctive properties. A single gamma ray can deposit enough energy in a cell to destroy it. In contrast, a single UV photon can, at most, mutate the DNA. Five million is a big factor.

fiber optics communications – needs high power

In Chapter 9, we explained that light is an extremely good way to send signals because of its high frequency. (Recall that Shannon's information theorem says that the bits per second is approximately equal to the frequency.) But now we can get an interesting result from quantum mechanics: high frequency isn't enough. We also need high power.

Here is why: a one milliwatt laser beam (typical for a laser pointer) has 10^{-3} joules/sec = 6×10^{15} eV per sec. Since each photon is 2.4 eV, this means that the light has a little over 2×10^{15} photons every second. The frequency for green light is 6×10^{14} Hz. So there are only about 3 photons per cycle, on average.

---

[4] For the experts only: there are exceptions, such as Compton scattering. But even in Compton scattering the gamma gives a substantial fraction of its energy on each scatter.

That is pretty low. Can you see why the communications would not work if the value were less than one photon per cycle? You can not send signals faster than the photon rate, and even three per cycle is low. The value of three is only an average number, and statistical calculations show that if the average is 3, then about 5% of the time there will be no photons at all in a given cycle, even when the cycle is supposed to have three. That means that if you send a bit for a "1", there is a 5% chance that it will show no photons, and be interpreted as a "0" bit. That is an error. The conclusion is that enough power must be used to avoid this “photon limit.”   To avoid high error rates, you need many more than one photon for each cycle.

So fiber optics communication requires high frequency (so there can be a large number of bits per second) and high power (so there can be many photons per bit).

Do photons really exist?

We’ve been talking about photons as if they are particles. Yet we know that they are electromagnetic waves. So how can we do that? Do photons really exist? Are they particles, waves, or both? Now we are discussing the heart of quantum physics. Don't expect simple answers to these simple questions. The answers are bound to be confusing. But let's give it a try.

Light behaves as a wave – except when it is emitted or absorbed. All the quantum features showed themselves only during these times. Of course, that is when we interact with them, so that is important. But in between – after emission and before absorption – the “photon” nature of light doesn’t seem to exist.

If that strikes you as weird, then I am glad. It is weird, and it still bothers many physicists. Let me illustrate what it means with a simple example, the soap bubble.

Recall that the colors of the soap bubble came about because some of the light wave bounced off the inner surface of the bubble, and some bounced off the front surface, and when these two waves came together, the waves interfered. Some colors (the ones that came out in phase with each other) were made stronger, and some (those that cancelled) were made weaker or nonexistent.

How does this interference fit in with the picture of photons? Let’s imagine that we turn down the intensity of the light until only one photon every minute is detected reflecting off the soap bubble. You might think that the photon was reflected off the outer surface of the bubble, or off the inner surface, but obviously it couldn’t have been reflected off both. So at very low levels of light, you would think that all the colors that arise from wave cancellation would disappear. You can't possibly have beats when only one photon is present! Right? Wrong.

The experiment has been done, not with soap bubbles, but with mirrors. In fact, it is not hard to do, and can be done by undergraduate physics majors in the upper division laboratory. The results are unambiguous. It is as if the photon split in half, and bounced off both surfaces. So the photon behaves like a wave, right up to the point where you detect it. Only then is the particle behavior evident.

In fact, the best way to think of light is as a wave that can be emitted or absorbed only in quanta – but that in between, it is a wave. It moves like a wave, diffracts like a wave, bends like a wave, and interferes like a wave. But it is not emitted and absorbed like a wave, but like a particle. This is, as I mentioned previously, the famous "**wave-particle duality**" of quantum mechanics that mystifies many people. But it mystifies them only because they think particles and waves are different things. I like to use the term "particle-wave" or "wave-particle" because real things have some of the properties of both.

We'll get back to these issues when we discuss the Heisenberg uncertainty principle. But for now let's return to practical consequences.

Semiconductor Electronics[5]

Essentially all modern electronics is based on the fact that electrons are waves. Their wave nature is very important when they flow through crystals known as semiconductors. (The word semiconductor comes from the fact that the material is not as conductive as a metal, yet it still conducts.) The most important semiconductors are silicon and germanium, often with small amounts of aluminum or phosphorus mixed into their crystals. Important applications of semiconductors include the microprocessor that runs your computer, laser diodes that play your compact disk players, and virtually all other modern electronics in your TV, your car, and even in screw-in fluorescent light bulbs.

The key feature of semiconductors, the one that makes it so important, is the fact that not all energies of electrons can flow. There is an energy gap (typically about an electron volt) just as there is in the hydrogen and helium atom. This energy gap is a result of the fact that electrons are waves. When electrons move through crystals, and their wavelength happens to match the crystal spacing, then there is a very strong reflectance that results in an energy gap. As with a hydrogen atom, the typical energy gap is a few electron volts.

Light emitting diodes and semiconductor lasers both take advantage of this energy gap to emit photons.

---

[5] Note for the experts: Most introductions to semiconductors emphasize the importance both of electrons and of objects call "holes" – bubbles in the electron sea, i.e. the absence of electrons. Holes behave much like positively-charged electrons. Although holes are important for semiconductor engineering, all of the key quantum-mechanics issues can be discussed without introducing them. They should be the first topic for students who want to go deeper into semiconductor physics than we do in this chapter.

Light emitting diodes (LEDs)

A light emitting diode (LED) is a semiconductor that emits light when a voltage is applied across it. Those little red lights that let you know your computer (or anything else) is on is usually an LED. The large TV displays used at stadiums and for some street displays are large arrays of red, blue, and green LEDs. LEDs light up your watch when you push the button. Many traffic lights are being replaced by LEDs because the LEDs are more efficient (they don't produce waste heat) and they don't burn out like tungsten filaments. In the near future, most flashlights will use LEDs instead of small tungsten bulbs. (Expensive flashlights already use them.) Infrared LEDs on your TV remote control send a burst of invisible light to the TV to tell it to turn on, off, or to change the channel.

An LED works in a simple way: an applied electric voltage gives an electron extra energy.[6] Because of the energy gap, it can't lose little bits of this energy, but only the entire amount, all at once. It does this by emitting a photon. The color of the photon is related to the energy gap by the Einstein formula $E = hf$. An LED with a small energy gap gives red light; an LED with a large energy gap gives blue light.

Look around and see how many LEDs you come in contact with every day. You may have to look at a traffic light up close to notice. In 2005 (when this is being written) many red traffic lights have LEDs, but the green ones still use incandescent lamps. Look on computers and stereo sets for the little light that comes on indicating that power is present.

Diode lasers

A diode laser is the kind that is used in supermarket scanners, in laser pointers, and in CD and DVD players to create the light that is reflected off the disk. It is very similar to an LED: it is a small semiconductor in which the electron is "excited" from its low energy to a higher one. The main difference between a LED and a diode laser is that the diode laser takes advantage of stimulated emission, i.e. the fact that one emitted photon can stimulate the emission of another photon. To achieve this, scientists had to find a semiconductor in which the photon would not be emitted spontaneously before it could be stimulated to emit.

Because the diode laser is a laser, it means that the photons that come out are all going in the same direction. This collimation is not quite as good as for a large size laser, but it is much better than in the LED, in which the light comes out in all directions. The collimation allows the light to form a very narrow beam that does not spread, and which

---

[6] Most books will describe in detail how energy is delivered to the electron. It usually involves a junction between two semiconductors that have different "doping" which leads to different energy levels. But the key reason that light is emitted is because of the energy gap.

can be focused to a very small spot. This is important for most of the applications mentioned above.

It is the diode laser that has really transformed the laser into an everyday device. Prior lasers looked like fluorescent light bulbs, but with the light coming out the end. They required special power supplies, and had very limited lifetimes.

Diodes – to turn AC into DC

One of the earliest (and still the simplest) semiconductor device is the diode rectifier, often called a **diode** for short. The reason this is so important is that almost all electronics requires DC (direct current) in which the electrons only flow one way. Yet, as we discussed in Chapter 6, the electricity that comes to our homes is AC (alternating current). A diode can turn AC to DC by letting through only the half of the current that is flowing in the right way. That's why it is called a rectifier – it allows it to flow only the "right" way.

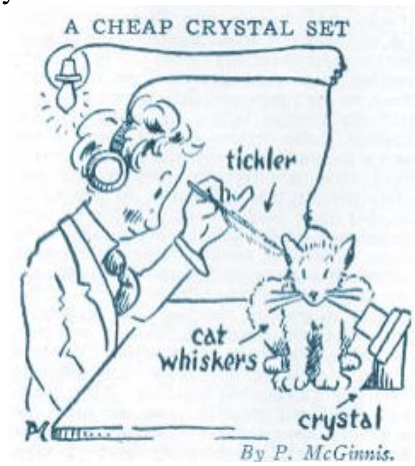
To make a diode you put two semiconductors that have different energy gaps into contact with each other.[7] As soon as that is done, electrons begin to flow from the one which has higher energy to the one that has lower energy. The flow finally stops when enough electric charge builds up to repel additional electrons. The same thing would happen with balls rolling down a hill: if they were charged, eventually the repulsion would keep other balls from rolling down.

The electrons that have accumulated create a strong electric field near the junction. This field prevents additional charges from flowing. If you weaken this field by applying an extra voltage (e.g. from a battery) then additional electrons will flow to rebuild it. But if you strengthen the field by applying an additional voltage in the same direction, then no current can flow. This is the basic idea that makes the semiconductor diode work. It lets current flow one way, but not the other. Put AC into it, and current will flow only half of the time – the half when it is going in the right direction. That's the way it turns AC into DC. (The magnitude of the DC will still vary with time, going up and down, but it will never go the other way. To smooth it out takes other electronics.)

Diodes have a long history. Like superconductors, they were used before they were understood. One of the earliest diodes ever used was the "cat's whisker" of the old crystal radio sets that amateurs (and even the author of this book) used to make as a hobby. A thin wire was delicately placed on a crystal and moved around until a spot was found that conducted electricity in just one direction. The wire was supposed to be as thin as a cat's whisker – so these were called cat's whisker diodes. Below is a cartoon

[7] For the experts only: The two materials are often both made out of silicon, but they have different impurities purposely mixed in, for example, aluminum in one and phosphorus in the other. This creates "donor levels" and it is the energy gap between these donor levels that plays the key role. Diffusion of charge carriers brings these donor levels to the same energy, and that diffusion creates the electric field at the junction.

from 1923 depicting an amateur radio builder trying to get his cat to move his whisker to just the right place on the crystal.



(found on the web site http://electronics.howstuffworks.com)

But real cat's whiskers were never used. It was a metaphor.

Transistor amplifiers

One of the most important electronic devices in stereos, TV, and almost everything else, is the *amplifier*. An amplifier takes in a weak signal (maybe from the reflected light from a CD, or from a weak electrical voltage from an antenna) and puts out a very strong signal, enough to drive a loudspeaker (if it is a sound signal) or a TV picture.

The word amplifier can be misleading, because an amplifier doesn't really take a weak signal and make it stronger – that might violate the conservation of energy. What it really does is take a weak signal and create a strong one that oscillates in the identical way. You could say it makes a clone of the original weak signal. The energy for that clone comes from a battery or other source of power (such as a wall plug electricity).

An amplifier is very similar, in some aspects, to a water valve. Think of the way that a water valve operates. You turn a small handle, and with very little energy, you control a huge flow of water.

The original electric amplifier used a device called a *vacuum tube*. Some old TV sets, and some stereos, still use vacuum tubes. (Some audiophiles claim tube sets sound better.) Vacuum tubes were based on the fact that electricity flowing through a vacuum could be changed by the small voltage.[8] (The vacuum was located inside a glass tube, and

[8] In the vacuum of the tube a tungsten filament heated a piece of metal called a cathode that emits electrons when it was hot. The electrons would flow through the vacuum to another piece of metal called an anode. In between was a grid of wires. Small voltages applied to this grid could make large changes in the current flowing through the grid. This design is still used in CRT (cathode-ray tubes) displays. The heated filiment made the vacuum tube hot, and used a lot of power. Tubes needed a lot of space for the electrons to flow through the vacuum, so they were large.

that's the origin of the term *vacuum tube*.) It was just like turning a valve to control water flow. In fact, what we call vacuum tubes in the US, are called electric *valves* in England.

One of the great revolutions in high technology took place when the vacuum tubes were replaced by transistor amplifiers. These produce the same effect - a small signal is reproduced into a large signal with exactly the same shape. But they don't involve vacuum or metal electron emitters. Rather they use semiconductors, the same materials we described in the previous section on diodes.

The original transistor amplifiers consist of two transistor diodes placed back to back. The middle part is made very thin. Very small voltages placed on the central part controls very large currents that flow from one side to the other. Again, as with diodes, the reason it works is that the three regions have energy gaps, and small changes in the voltages can affect the relative energies of the gaps. It is analogous to having a large dam, with a huge reservoir. Small changes in the height of the dam can cause huge changes in the amount of water that flows over.

More modern transistor amplifiers use a different design called a Mosfet. I'll describe the Mosfet in an optional section below.

All radios used to use vacuum tubes as amplifiers. In the 1960s, transistors began to replace the tubes. The new radios were called "transistor radios." Eventually people started called them simply "transistors."[9]

Transistors are far more reliable, much faster, take less power, and generate less waste heat then vacuum tubes. They were invented in 1947.[10] In the 1960s, a good (and expensive) portable radio contained, typically, 8 transistors. But the size and cost of transistors has continued to decrease. An important breakthrough came when engineers figured out how to put many transistors on a single chip of silicon – creating a device called the "integrated circuit." The Nobel Prize in 2000 was awarded for this invention.

The integrated circuit really made Moore's Law start to operate (see Chapter 5). As the transistors were made smaller, the complexity of circuits could grow. The first full microprocessor (computer that has all of its complex circuitry on one chip of silicon) was created in 1971. Now we have over ten million of the quantum devices called transistors

[9] Some older folk are still bothered by the dropping of the word "radio". But they forget that a radio was once called a "radio wave receiver". It was quite different from a "radio wave transmitter." So they shortened it to simply radio.

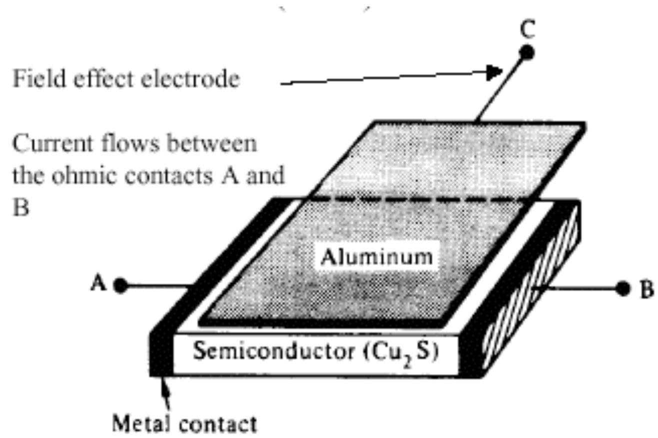
[10] In 1956 the Nobel Prize in physics was awarded to W. Schockley, J. Bardeen, and W. Brattain of Bell Telephone Laboratories for their invention of the transistor.

on a single chip of silicon. That is a typical number for the biggest chip in your home computer. Such numbers were inconceiveable just a few years ago.

Optional: How a MOSFET transistor works

Because of the importance of transistors, there has been a huge development of the technology, and many different kinds have been created. I'll describe one of the simpler, but also one of the commoner types of transistors – an FET (which stands for "field effect transistor"). A variation of this is called a MOSFET (for metal-oxide-semiconductor FET). If you look up these terms on the web, you'll get lots of references.

In a FET, a current normally flows along a narrow semiconductor "channel." In the diagram below, the semiconductor is a compound of copper (Cu) and sulfur (S).



The current normally flows from A to B. But a small voltage on the aluminum cover C can make big changes in that current. Here's why it works: Remember that when we put two semiconductors together, if they had different energy gaps then the electrons tended to flow to the one with lower energy. The same thing happens if you put some positive charge next to the semiconductor, on the piece of aluminum labeled "C" in the diagram. When that is done, then electrons are drawn to the metal and they cross over; as a result, there are fewer electrons in the semiconductor, and that decreases the current that flows from A to B. In other words, a small voltage on the aluminum *depletes* the semiconductor of its conducting electrons, and that affects the current flowing through that semiconductor.

Another diagram of a semiconductor switch in use is shown below.

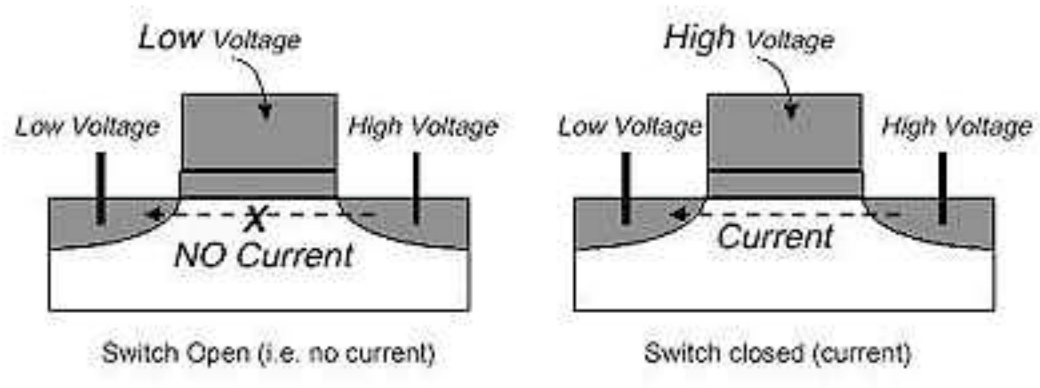


image borrowed from http://www.sysopt.com/articles/soi/index2.html

Computer circuits

All "computation" done in a computer is done with special amplifiers called "switches". Linked switches can be used to add numbers, multiply them, divide them, or make logical conclusions. But a computer switch is nothing more than a transistor amplifier. It is called a "switch" because it is designed to have only two positions: all the way on, or all the way off.

Computers have become very complex. The original computers had thousands of vacuum tubes working as switches. These computers were large, used a lot of power, and needed frequent maintence (replacement of tubes as they burned out). Modern computers use transistors for their switches, and often contain more than 10 million of these transistors. The transistors are made very small (a few microns in size) so they require very little power, and can be put together into very small chips. The small size is important for speed, since the velocity that signals travel down wires is always less than the speed of light. In one typical computer cycle (one nanosecond; a billionth of a second) light can travel only 30 centimeters – one foot. So to be able to exchange information with other parts of the computer, the whole thing must be small.

optional – a little more about logic switches

Let A and B be statements that could be true or false. Suppose A implies B, i.e. if A is true, that implies that B is true too. (Here are two possible statements: A: John is Mary's father. B: John is male.) In computers this logic is performed with a special amplifier called a switch: if the switch A is turned on (i.e. A is true) than that switch turns on the switch B (showing that B is also true). Some switches require two inputs to be turned on; others will be turned on if either of two other switches is turned on.

All computation done in computers is done by connection of switches such as this. One of the most fascinating theorems of computer science is that every known computation (those doable using logic or math) can be accomplished by using such switches, in conjunction with other devices that work as the memory.

The physics of superconductors

When we talked about superconductors, we never explained why the electrons can move through the superconducting metal with no loss of energy. In fact, superconductivity was discovered by Omnes in 1911 (he was awarded the Nobel Prize in 1913) and yet the phenomenon was not understood by him or anyone else for many decades. For much of the 20^{th} century it was the outstanding failure of the quantum theory that nobody could figure out why superconductors were superconductors!

The reason did turn out to be quantum mechanical, and just as with spectral fingerprints and semiconductors, the answer was the existence of an energy gap. Just as in the other materials, the behavior of superconductors comes about because electrons are waves, and in certain crystals this can lead to an energy gap – certain energies that an electron can not have when it travels through these crystals. The energy gap is only 0.001 eV – but that is enough to give zero resistance.

Now here is the key reason for superconductors: the flowing electrons all have a low amount of energy. In ordinary metals, such as copper wires, they would collide with impurities in the metal, and gain or lose energy. But in a superconductor, the existence of an energy gap means that the energy of the electron cannot change by small amounts. They can't change energy gradually, even if from small collsions. In the almost perverse logic of quantum mechanics, that means that can't collide with impurities – because losing energy is impossible! So the electrons will continue to flow without energy loss. That's what superconductivity is.

The strange concept in the last paragraph is not completely new. Physics often states that behavior that contradicts the conservation of energy cannot happen. What seems most weird in superconductors is that the impurities became invisible to the electrons, because if they bounced off them they would have to have illegal (impossible) energies.

Not all metals become superconductors at low temperatures. We now understand that for a metal to become superconductor, an interesting thing must happen: the electrons have to move in pairs. This happens when slow moving electrons pull the positive charges of the metal close to them, and that distortion of the metal tends to attract another electron. The net result is that electrons attract other electrons. The electrons never get very close, and it doesn't take very much energy to break them apart. That's why this happens only at low temperatures. The two electrons are called "**Cooper pairs**" after the name of the physicist who first predicted their existence.

The full quantum theory of superconductivity was worked out in 1957 by John Bardeen, Leon Cooper, and Robert Schrieffer. They were awarded the Nobel Prize for this work in 1972. From their theory they could predict which metals would become superconductors and which won't, and what temperatures it happens. Some of the strangest results of their theory is that hydrogen can become a superconductor at high pressure. That means that the core of the planet Jupiter may be a superconductor, as well as the stars known as pulsars. These predictions have not yet been verified.

Despite the fact that we understand superconductivity at low temperatures, once again, superconductivity has a mystery. This time, it is high-temperature superconductors, the ones that become superconducting at temperatures up to 150 K. The BCS theory does not predict their existence, and nobody has been able to figure out why these compounds become superconductors at such high temperatures. We do know that the flow involves Cooper pairs. But nobody has a good theory to predict which materials will become high temperature superconductors, or how high the superconducting temperature can go. It a material is found that is superconducting at room temperature, then we will see a new technology revolution that will be bigger than the one involving transistors. Energy transport will become very easy; energy loss in computers will become very small; we may switch from AC in our homes to DC; …

The electron microscope

Objects that are less than a micron in size cannot be resolved with ordinary light because you cannot focus a beam on a spot smaller than the beam's wavelength. If you want to look at something smaller than that, you need a wave with a shorter wavelength. X-rays are sometimes used, but x-rays tend to go right through objects, especially objects that are only a few microns thick. A more widely used option: use electrons. Electron beams with an energy of 50 keV have a wavelength smaller than the size of atoms[11].

There are several kinds of electron microscope, but the most interesting one is the "scanning electron microscope" or SEM. In an SEM, a beam of electrons is scanned across the object, much as a beam of electrons scans on the surface of a material, and the number which bounce off in a particular direction is measured. This number is called the "brighness", and based on it, an image is created in a computer. These images look remarkably like ordinary photos. That's probably because shadows make it look realistic. (The shadows are there because electron beams hitting the back side of an object reflect away from the detector, and fewer of them are collected.) Below is an SEM image of the claw of a spider.

[11] Optional: Since electrons have mass, according to the theory of quantum physics (not all of which we have discussed) the wavelength L has to be calculated from the "deBroglie equation" $L = h/(mv)$. The momentum mv can be calculated if you know the energy $E = 1/2\ m\ v^2 = (mv)^2/(2m)$.

Image of a spider claw (black widow) taken with a scanning electron microscope
(from http://www.mos.org/sln/sem/widow.html)

Look on the web for more SEM images. You'll find things that are truly amazing. Search for images that have the word SEM or electron microscope associated with them.

Are ordinary waves quantized?

Light waves are quantized. Electron waves are quantized. Do you think that water waves are also quantized? Are there only certain allowed energies that a water wave can have? What do you think?

The surprising answer is: yes. How can that be? Why don't we notice?

Let's look at a typical water wave. It might have a frequency of $f = 1$ cycle per second, (Such a wave has a wavelength of about 1.6 meters.) The energy will be quantized by the Einstein equation $E = h f = 6.6 \times 10^{-34}$ joules. That is tiny. If a wave hits you, and delivers an energy of one joule, then that consists of $1/(6.6 \times 10^{-34}) = 1.5 \times 10^{33}$ quanta. With that many wave quanta, the fact that the energy is quantized is impossible to notice. This is a case in which a "quantum leap" is very tiny indeed.

In fact, the quanta are so small that they have never been observed for water waves. They don't even have a name. (Waterons? Hydrolons?) But according to quantum theory, they are there.

A similar quantization happens for all those waves that appear to us as real waves: sound waves, low frequency radio and TV waves, rope and slinky waves. They are quantized, but in the limit of low frequencies, the quanta are so small that we would never notice.

Is the Earth's orbit quantized?

Yes. Why did we never notice? The reason is that the gaps between the energy levels is tiny. A calculation using quantum mechanics shows that the energy gap between different orbits of the Earth is about 0.001 eV – that is, it is a thousand times smaller than the typical energy gap in a single atom. Compared to the energy in the motion of the Earth, this "gap" is only one part in 10^{55}. That's why it can't be observed.

Is the entire Earth really a wave? Again, the answer is yes, but it is a complicated wave, wwith many parts. Speech is also complicated, and waves in the ocean during a storm are complicated; light wave from the page that you are reading are complicated. They consist of waves of many wavelengths all mixed together. Because of this complexity, it turns out not be be very useful to think of the Earth as a wave, even though it is. In the atom, where the energy gap is comparable to the electron energy, then the wave concept is not only useful, but essential.

Uncertainty in quantum physics

The fact that a particle is a wave implies that its position is uncertain. Waves don't exist at a single location; they are spread out. This is the heart of the famous "uncertainty principle."

But let's now think about the energy. According to the Einstein equation, the energy is the frequency.   But what is the "frequency" of a wavepacket? In the middle of the packet, the oscillations seem pretty regular, but near the ends, they die out. Does that mean that the frequency changes? Remarkably, the answer is yes. This is a bit tricky to understand, but it is true. Short bursts of sound don't have the pure single frequency of sound that a long tone will have. A short burst of light will spread in a prism, because the sharp changes in intensity cause other frequencies to be present in the burst. Mathematically, there is a simple relationship: the bandwidth (the range of frequencies present) is given by B = 1/T, where T is the time duration of the wavepacket.

But – if several frequencies are present in a wavepack, that implies that there are several energies present. A wave packet does NOT have definite well-defined energy!

That idea is so disconcerting, that Albert Einstein was never able to accept it. He worked for years trying to show that quantum mechanics was not the true theory of matter, but was only an approximation. He believed the real, underlying theory, would bring back definite energies and definite positions. But he never succeeded. You will sometimes still see his name associated with arguments against quantum mechnics. But every testable theory proposed to replace it, so far, has been proven wrong by experiments.

measurement of the wave

Suppose a very broad wavepacket of a photon hits the CCD (the light sensitive detector) in a digital camera. You can even assume that the width of the wavepacket is larger than the entire camera. Because the packet is so large, does it give a signal in every pixel of the CCD? No. It knocks an electron out of only one pixel – because it has only enough energy to affect one.

Which one? Quantum physics supplies a curious answer. The electron knocked out will be random, but only in the region where the wavepacket hits. The places where the wavepacket is stronger (greater amplitude) have a higher probability. In fact, according to the math of quantum physics, the probability is proportional to the square of the amplitude.

So quantum mechanics is inherently a random theory. There is no way to predict where the electron will be knocked out.

Some people like to say that the position of the knocked out electron represents the "true" position of where the photon "really was." But that's wrong. The wave is spread out over space, and the photon is not hidden within this range. We know that because of soap bubble – the photon can bounce off both the front and the back of the soap bubble, because it is a wave, not a localized particle.

We came across this before, when we discussed radioactive decay. Two identical nuclei of tritium will emit their beta particles at different times – despite the fact that the nuclei are identical. Quantum physics can give only the probability that the beta will be emitted.

optional: the precise statement of the uncertainty principle

You can make the position of a particle certain by making a very small wave packet. But in quantum mechanics, such a wave is made up of many different waves traveling at different velocities. Yet only one particle is present. So once you detect the particle, all the other waves have to suddenly disappear. This sudden disappearance has a special name in quantum physics: it is called "the collapse of the wave function."

So although the position of the electron may be well known, the velocity is uncertain. That means its energy is uncertain too. These relations are at the heart of the famous *Heisenberg uncertainty principle*. It says if you create an electron with very well-defined position, so that it is known to an accuracy Δx (which can be smaller than an atom, or larger than the sun), then the velocity is uncertain by an amount at least equal to:

$$\Delta v = h/(2\pi\, m\, \Delta x)$$

In this equation, m is the mass of the electron, and h is Planck's constant. So if you improve your knowledge of the position (i.e. make the wave pack, and Δx, small by passing the wave through a small hole), that makes the uncertainty in velocity greater.

A similar thing happens with light. Let's talk about knowing the position of the photon in the x-direction, for a beam of light that is traveling in the y-direction. To determine the x position, you let the light wave pass through an opening of width D. But in doing that, you make the wave spread out, from diffraction. That means that the velocity in the x-direction is no longer certain; part of the wave is moving to the left, and part to the right. When the photon is detected, it could be moving (at least in part) sideways. It turns out that the blurring equation that we gave in Chapter 8 is also the Heisenberg uncertainty relation – but in the special form that is appropriate for light.

is energy still conserved?

If energy is uncertain, does that mean we can lose some? Good question. Remarkably, the answer is no. (Note how often I use the term "remarkable" or "weird" or "strange" in the discussion of quantum physics. That represents the fact that most physicists who think deeply about these issues still find them peculiar.)

Suppose we have a definite amount of energy, and give it to two electrons. Assume each electron is a wave packet, so its energy is uncertain. The electrons move apart. Each one has uncertain energy. Suppose each travels a mile, or a million miles, in opposite directions. Finally, we detect each electron. The energy we get could turn out to be anything within the uncertainty. But when we add together the two energies we measure, the sum will be the same as we started with. Energy is conserved.

Does that mean that the energy really had a certain energy, but we just didn't know what it is? No, that possibility is called "hidden variable theory" (look it up on the web). It makes other predictions that have been tested and found to be wrong. There have been many versions of hidden variable theories, but that have been tested have been shown wrong.

Maybe some day a new hidden variable theory will be invented that works, and the Einstein's spirit will smile.

tunneling

Tunneling is one of the more famous phenomena in quantum physics. It says that particles can travel to places where they appear to violate the conservation of energy. Tunneling is a consequence of the uncertainty principle, in particular, the fact that for a wave packet the energy of a wave is uncertain. The name "tunneling" comes about because, in effect, a particle can go from one side of a hill to the other, even though it doesn't have sufficient energy to get over the hill.

Tunneling is relatively easily to calculate when you know the height and width of the hill. We teach junior physics majors how to calculate it. Like other things in quantum mechanics, calculations gives probabilistic results. You can't say for sure that something will tunnel, but you can calculate the probability that it will tunnel in a given time. We won't do the calculation here, but instead will discuss the consequences of tunneling, and the practical application in the tunnel diode.

alpha radiation

Remember alpha particle radioactivity, from Chapter 4? It turns out that this kind of radioactivity occurs because of tunneling. The alpha particle is inside the nucleus prior to the decay, but there are forces that prevent it from coming out. According to ordinary physical laws, it doesn't have enough energy to overcome the attraction of the nuclear force. But, thanks to the uncertainty principle, there is some chance that it will tunnel out anyway. Its energy is uncertain, and therefore there is some small chance at any moment that it will have enough energy to escape. Because nobody can calculate when it will come out, but just the likelihood that it will come out in any time period, the decays occur randomly.

(It is worth pointing out that not all radioactivity is due to tunneling. In beta decay, the electron and neutrino are both created at the time of their emission. They are like the sound waves that you create when you speak; they didn't exist until they are created at the moment of decay, but when they are created, they can carry away energy. Likewise, x-ray and gamma ray radioactivity is not an example of tunneling.)

So every time you see an alpha decay, you know that tunneling has taken place. Energy conservation was violated, but only for a short period of time. Once the alpha particle is out, the energy it has is identical to the energy it had inside the nucleus. We never actually see it violate energy conservation. We just calculate that it must have done so, but only for a very short period of time. So, in the end, energy is conserved. There is no more energy than there was before the decay. Somehow the alpha particle snuck through. We say that it tunneled.

tunnel diodes

One of the more practical uses of tunneling is a kind of semiconductor amplifier known as the tunnel diode. The tunnel diode works very much like an ordinary diode, with two semiconductors with different energy gaps put in contact. Electrons move across the junction because of the lower energy gap on the other side, and it doing so they create an electric field that repels other electrons. So far, that is the same an ordinary semiconductor diode.

In a tunnel diode, that region is made very thin, and a battery voltage placed across the junction gives electrons almost enough energy to get across – but not quite enough. However, thanks to tunneling, some do leak through. In a tunnel diode, impurities are added to the semiconductor to make sure that there are lot of electrons available for this leakage. The current that gets through is called the tunnel current.

The number that tunnel depends on the electron energy. So, if a varying voltage is added to the battery voltage, then the energy of the electrons changes, and that makes tunneling easier or harder. The effect is very strong, so a little bit of voltage change can make a big current change. That is what makes the device into an amplifier. That's what an amplifier is: something where a small change in voltage can make a large change in current.

Tunnel diodes also respond very rapidly to changes in current, faster than do FETs. By making the voltage changes large enough, the current essentially switches on and off. Because tunneling is such a fast effect, tunnel diodes are among the fastest electronic switches we can make, and thanks to that, they are used when high speed switching is important, for example, when routing signals along the internet.

scanning tunneling microscopes (STMs)

One of the newest and most powerful microscopes, one that allows us to detect the positions of individual atoms, is the scanning-tunneling microscope, or STM. Recall the image of "IBM" shown at the beginning of chapter 2? That image was taken with an STM.

An STM consists of a small sharp needle point with a electric charge on it. The needle point is brought very close to the surface that will be examined, but it doesn't quite touch. Then the point scanned (that's the S part of STM) just above the surface. It moves back and forth, eventually over the entire surface. It picks us small electric current through tunneling (we'll come back to that). The amount of current is then recorded. Finally, the computer puts together a map of the current. More current might be made whiter, and less current darker. The resulting map is the image.

The process is similar to that of a blind person feeling a statue. His fingers move over it, and after they have covered the whole surface, he knows that statue as well (maybe better) as someone who looks at it from every angle. (The sighted person may know more about the color, if there is any, but the blind person knows more about the texture.)

The key to the STM is the tunning process. Normally electrons cannot leave the surface of the needle tip because they don't have enough energy. However, if the tip is brought very close to another atom, then the distance is so small that tunneling occurs. The smaller the distance, the larger the tunneling. So measuring the current flowing gives the shape (ups and down) on the surface.

Care must be taken so that the needle never actually touches the surface. So an STM can be operated in a different mode. The needle tip is attached to a piezoelectric crystal – a crystal whose thickness can be adjusted to very high precision by applying a voltage across it (the crystal voltage). As this point is brought closer to the surface (by adjusting the crystal voltage on the piezoelectric), electrons begin to tunnel across from the tip to the surface. When a certain amount current flows, the tip is brought no closer. Then the tip is moved across the surface.

Here is the tricky part: the position of the tip is adjusted as much as necessary, in order to keep the tip current constant. That means that the tip is being moved up and down (by changes in the crystal thickness) to keep it a constant distance above the surface. The crystal thickness needed to do this is recorded, and it becomes a record of the height of the surface at every location. That record can then be made into a map of the surface height.

The STM was also used to place the xenon atoms in the IBM photo. By moving the tip extra close, and adjusting the voltage, the atoms could be picked up and put down. So for that manipulation, the needle tip actually does touch some atoms.

STMs are the best way to get images of atoms, and their positions. Right now, their main use is to study the properties of surfaces, and the way atoms are bound to those surfaces. In the near future, STMs may be used to scan across DNA molecules to read the genetic code. Some people think they may be used to store information by adjusting the positions of individual atoms, but I am guessing that that is unlikely, at least for the next ten years.

tunneling in the Sun

As we described in Chapter 5, the Sun is powered by nuclear fusion. At high enough temperatures, protons, deuterons, and other positively charged nuclei have enough kinetic energy to overcome their electric repulsion, so that they get close enough that the nuclear force brings them together and they fuse.

However, calculations show that the Sun is not hot enough to bring the nuclei that close. Their thermal energy brings them near each other, but not enough to fuse. Yet they fuse anyway. The reason is tunneling. Once the nuclei get close, there is a high probability that they can tunnel right through the barrier of repulsion (it is completely analogous to pushing a weight up a hill) and get close enough for the fusion to take place. In that sense, essentially all of the energy we have on Earth is produced using tunneling. The same process takes place in all stars.

Tunneling is also important in nuclear fission. Calculations show that the forces holding the two fission fragments together are quite strong, too strong to ever let them break apart. But because the fission fragments behave as quantum-mechanical waves, they can overcome this energy deficiency if they do it fast enough. Without tunneling, we would not have fission and its applications (reactors and bombs).

quantum computers

Unlike most of the other technologies described in this book, real quantum computers don't yet exist. The only kind that have been made perform extremely simple operations, such as adding the number 1 to the number 1. Nobody knows whether they will ever prove practical. Yet there is a great deal of interest in them, and so they are worth mentioning.

All computers use quantum mechanics, in that the energies of the electron flow in semiconductors is quantized. The random memory of a computer is based on the storage of electrons on small pieces of metal on the chip surface. Electric charge is quantized, that is, it is always present in some multiple of the electron charge. But despite all these ways that ordinary computers are "quantized," none of them are what we mean by quantum computers.

In ordinary computers, charge is stored, and it flows though switches. Every computation consists of changing the stored charge by regulating the flow of electric current. But in a quantum computer, the idea is fundamentally different. All manipulations are done with the electron wave rather than with the current. This can be done by changing the wave with an electric field or some other external force. No particle is measured or stored until the computation is all finished.

By working with the waves themselves, a much greater amount of information can be stored, a very large number of computations can be carried out simultaneously using very simple circuits, and much less energy can be used in computation. In a sense, the quantum computer will take advantage of the uncertainty principle. By being careful not to detect the electron, the spread out wave can carry more information than the simple presence or absence of the electron. Each electron can, in principle, carry the equivalent many bits of information. They are called quantum bits -- "qubits" for short.

At least that is the theory. Nobody knows whether quantum computing will ever prove practical for really large and difficult computation. Part of the problem is, of course, that we already have pretty good computers for most pretty hard problems, so the quantum computer has to make a lot of progress before they would be put to use for any practical purpose. For the latest, do a web search on quantum computing.

It is important to recognize that most new technologies never become practical in the way that people who try to look far ahead, sometimes called "futurists," speculate. For example, in the 1920s, and every decade since, futurists have predicted that ordinary people would soon be driving their own airplanes instead of cars. They predicted that this was such a certainty that it surely would happen by the 1940s. Yet, it hasn't happened yet. In the 1940s, futurists predicted that we would have robots helping us in our homes, certainly by the 1960s. Yet that hasn't happened. Other things (like laptop computers) have happened. The future is hard to predict. Quantum computing has lots of obstacles before it can become practical, and some of them are fundamental (such as keeping noise out of the computation). They may never become useful. But they may.

Quick review

Electrons, protons, and all other particles are quantum waves, in the same sense that a light wave is a wave. Their particle properties refer to the wave they behave when they are detected or measured. The wave nature is most evident in the way these object move from one place to another. One important consequence of their wave nature is the quantization of energy levels, both in atoms and in crystals.

Lasers depend on the fact that the presence of a photon will trigger "stimulated emission" of another photon, as the electron that had that energy changes energy, and this results in a chain reaction of photons. The emitted photons have the same frequency as the one that stimulated it, as well as the same direction. That means that a laser beam spreads very little, and it can be focused to a small spot, and that means that the energy it delivers can be strongly concentrated. That feature allows it to be used for laser cleaning and for surgery. Laser applications that make use of one or more of its properties include CD and DVD sensors, supermarket scanners, weapons, and laser printers.

The relation between the frequency of the light and the energy of the photon is given by the Einstein relation, $E = h\,f$. X-rays and gamma rays have very large values of photon energy; that's why they seem to be more like particles than do other electromagnetic waves. Visible light photons hitting a surface can give this much energy to an electron, a process called the photoelectric effect. That can eject it, or give it enough energy that it can be conducted away. The photoelectric effect is used in solar cells, digital cameras, Xerox machines, and image intensifiers.

Semiconductors such as silicon also have their special properties because of quantum mechanics. Put two different semiconductors together, and if their energy levels are different, then some electrons will flow from one to the other. This can be used to make a semiconductor diode (which lets current flow only one direction, thereby converting AC to DC), or a transistor amplifier. Virtually all modern electronics is based on diodes and transistors. A switch, the basic computation element of computers, is a version of the transistor. Integrated circuits consist of thousands to millions of transistors on one piece of semiconductor.

Superconductors work because at low temperature the electrons form Cooper pairs, and the motion of these pairs has an energy gap that prevents the pairs from losing small amounts of energy from collisions. As a result, they lose none.

Electron microscopes work by focusing a very fine beam on the object that is being observed. They can see things much smaller than can visible light microscopes. That's because the wavelength of an electron is smaller than the wavelength of light.

All waves are quantized, but for low frequency waves (water waves, radio waves, sound waves) the quantum of energy is so small that it is impossible to detect the tiny quantum leaps.

The Heisenberg uncertainty principle is a consequence of the fact that electrons and other "particles" behave as waves. Not only is location uncertain, but so is velocity and energy. Uncertainty in energy allows electrons to "tunnel" through regions that appear to violate the conservation of energy. Such tunneling is responsible for alpha radioactivity. Tunneling finds practical applications in the tunnel diode and in the scanning-tunneling microscope.

Quantum computers make use of the wave nature of particles to store information in qubits. Nobody knows if they will prove practical.

Internet research topics

Find images taken with electron microscopes. For each one, either find out what the magnification is, or estimate it by finding the size of the object imaged. Find some objects that have images both from ordinary visible light and from electron microscopy. What can be learned best from each kind of image?

Find commercial applications of lasers other than the ones I described in this chapter.

What is the current status of lasers as weapons? Are any being tested? Are any being deployed? Can you find discussion of their potential use for anti-ballistic-missile systems (ABMs)?

Find applications of spectral lines. Which of the systems you find can be used remotely (e.g. in the open atmosphere) and which require laboratory measurements? Find uses for environmental measurement, industrial measurement, and purely scientific (e.g. determining the composition of stars).

Find out what you can about high-temperature superconductors. Do they involve Cooper pairs? Have they estimated the size of the energy gap? What is the highest temperature that people have obtained? Are scientists optimistic about reaching higher temperatures?

Look up "crystal radio" and "cat's whisker diode" on the web and see what you can learn about the early days of radio reception, including the use of vacuum tubes as diodes.

What is the current status of quantum computing? What is the most complex calculation that anyone has done? What are the problems that are too difficult for ordinary computers that might be solvable if quantum computers become viable?

Essay topics

Lasers have several properties that make them special. Describe what these properties are. For each special property, describe practical applications that take advantage of it.

Quantum physics has properties that seem very strange to someone who doesn't recognize that particles have wave-like behavior. What behavior of electrons would be impossible to understand based on classical (non-quantum) physics? What behavior of light would be impossible to understand based on the classical theory of light (i.e. light as a wave, not a quantum wave)?

Describe phenomena that depend on the presence of an energy gap.

What methods can be used to identify gases, for example, to tell whether a gas is hydrogen or helium or a mixture? Describe the principles upon which the method is based.

What phenomena can be understood as a consequence of tunneling? What practical applications are there to this behavior?

Short questions

Stimulated emission is important for
( ) integrated circuits
( ) superconductors
( ) LEDs
( ) lasers

Transistors use semiconductors with different
( ) frequencies
( ) energy gaps
( ) densities
( ) wavelengths

Compact disk readers use
( ) x-rays
( ) lasers
( ) LEDs
( ) spectra

To tell hydrogen from helium, look at
( ) their x-ray emission
( ) their visible spectrum
( ) the photoelectric effect
( ) their amplification

Quanta are not observed from water waves because
( ) such waves are not quantized
( ) the energy of the quanta are too small
( ) water atoms are too small
( ) water atoms are too large

Xerox machines make use of
( ) the photoelectric effect
( ) Cooper pairs
( ) stimulated emission
( ) a chain reaction

Spectral lines occur because
( ) electron energies are quantized